心理科学进展 2021, Vol. 29, No. 8, 1508–1520
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2021.01508

道德困境研究的范式沿革及其理论价值*

刘传军[1,2,3] 廖江群[3]

([1] 四川大学公共管理学院社会学与心理学系; [2] 四川大学心理所, 成都 610065)
([3] 清华大学社会科学学院心理学系, 北京 100084)

摘 要 通过评述道德困境研究范式的发展过程, 系统阐释了经典两难法、加工分离法、CNI 模型法和 CAN 算法的优缺点和理论价值。后来的研究范式均在一定程度上克服了之前研究范式的局限。加工分离法克服了经典两难法的加工纯粹性假设等局限, CNI 模型法在加工分离法基础上进一步分离了道德困境决策的多种心理过程, CAN 算法则修正了 CNI 模型法的序列加工的不恰当预设。研究范式的沿革启示研究者综合应用新方法来解决研究争议和重新审视以往道德理论, 合理应用新方法来探索其他具有潜在冲突性的研究议题。总之, 本文为道德困境及相关研究提供了方法学参考。

关键词 道德困境, 加工分离, CNI 模型, CAN 算法, 道德决策

分类号 B849: C91

1 引言

随着无人驾驶汽车的到来, 道德困境问题再次引起热议。以往那些假设困境在现实中发生的可能性并不高, 但无人驾驶时代的到来, 则将道德困境问题实实在在地呈现在人们面前(Bonnefon et al., 2016; Greene, 2016; Young & Monroe, 2019)。当无人车在行驶过程中遇到紧急情况时, 如果按原定路线直接开过去将撞死 5 个甚至更多行人, 如果通过改变方向撞向路牙则会牺牲掉汽车及其乘客(人数相比行人更少)。在这种紧急情况之下, 无人车应该如何反应? 近期研究发现, 人们倾向于将无人车的程序设置为牺牲掉汽车及其乘客, 但并不愿意乘坐或购买这样的汽车(Greene, 2016; Young & Monroe, 2019)。

与无人驾驶类似的两难困境, 最早出现在道德领域。1967 年, 福特提出了著名的电车难题(Foot, 1967), 标志着道德困境研究范式的形成。大约 10 年后, 汤姆森又提出了另一个经典道德困境——天桥难题(Thomson, 1976)。电车难题中, 决策者如果选择扳道岔, 则会牺牲一个无辜者并拯救 5 个人, 但如果不扳则会牺牲 5 个人。天桥难题中, 决策者如果从天桥上推倒一个体型硕大的陌生人来阻挡失控的电车, 则会牺牲这个陌生人并拯救 5 个人, 但如果不推则会牺牲 5 个人。对于牺牲 1 人(或少数人)而拯救 5 人(或多数人)的功利性提议, 表示赞同则被认为是功利主义的道德判断, 因为这种判断是基于结果最大化来考虑的, 符合最大化人类福祉的功利主义原则(Bentham, 1996; Mill, 1872); 而表示拒绝则被认为是义务论的道德判断, 因为这种判断源于行为本身是否符合道德规范, 反映了符合义务论原则对行为规范合理性的内在要求(Kant & Gregor, 1997)。

电车难题中无辜者的死亡是由电车直接导致而非决策者亲手造成, 因此被称为非个人道德困境(impersonal dilemma)。相反, 天桥难题中, 无辜者的死亡是由于决策者直接推倒从而亲手造成其死亡, 因此被称为个人道德困境(personal dilemma)。这两类困境在后续的道德研究中被奉为经典, 许多研究探讨了这两种困境及其变式中人们的决策偏好及其成因与后果(Fumagalli et al., 2010; Gold et al., 2014; Graham et al., 2016; Greene et al., 2001; Valdesolo & Desteno, 2006), 比如有研究者发现男性在个人道德两难中更加功利主义, 但在非个人

收稿日期: 2020-11-27
* 国家社会科学基金项目(18BSH114); 清华大学自主科研计划(2017THZWYY11)。
通信作者: 廖江群, E-mail: liaojq@tsinghua.edu.cn

道德两难中无此性别差异(Fumagalli et al., 2010)。

随着研究推进, 许多道德困境及其变式被开发出来, 研究者所使用的情境材料各异, 缺乏标准化的研究方法(Christensen & Gomila, 2012)。厘清道德困境研究中的实证方法论范式对于道德困境以及类似的社会困境的研究, 均具有重要意义。本文将对目前主流使用的经典两难法、加工分离法、CNI (Consequence, Norm and generalized Inaction preference, CNI)模型法以及新近的 CAN (Consequence sensitivity, overall Action/inaction preferences and Norm sensitivity, CAN)算法进行综合分析并论述其方法学价值。

2　经典两难法

经典两难法主要是指电车难题、天桥难题及其变式, 通常呈现一个假定情境, 在这个情境中需要牺牲 1 个(或少数)无辜者来拯救 5 个人(或多数人), 决策者被提问这样做在道德上是否可接受以及是否愿意执行这样的行为提议等。研究者运用经典两难法来研究普通人的道德判断, 并提出了著名的道德双加工理论。心理学家 Greene 在早期研究中与道德哲学家 Haidt 一起, 使用经典两难法研究发现, 情绪在影响道德判断方面起着重要作用, 特别是对个体的义务论倾向具有预测力(Greene & Haidt, 2002; Greene et al., 2001)。而后来, 他也发现理性思考对道德判断同样起着非常重要的作用, 特别是对功利主义倾向具有预测力(Greene, 2007; Greene et al., 2008)。于是, Greene 提出了道德双加工理论(Greene, 2009), 对道德研究起着深远影响, 后续许多研究都在该理论框架下展开(Gubbins & Byrne, 2014; Lucas & Livingston, 2014; Patil et al., 2021; Skulmowski et al., 2014; Youssef et al., 2012; 喻丰 等, 2011)。

后续的实证研究广泛使用了经典两难法, 并使用道德双加工理论来解释研究发现。但是, 这些研究中出现了许多不合常理、分歧甚至矛盾的结果。首先, 在功利主义反应倾向方面, 有研究发现决策者的精神病性、马基雅维里主义和生活无意义感越强, 其赞成牺牲少数人来拯救多数人的倾向更高(Bartels & Pizarro, 2011), 特别是精神病性越强, 其赞成功利性提议的概率越高, 在后续研究中多次被证实(Seara-Cardoso et al., 2013; Tassy et al., 2013; Yu & Tang, 2013)。对行为结果的利弊权衡一直被认为是道德理性的表现, 而这却与精神病性具有正相关关系, 这意味着相较于常人, 精神病性越强或者精神病人更加理性和关心人类福祉的最大化, 或者更加理性和关心最大化人类福祉的人可能会有更强的精神病性倾向, 这在理论和常识上都是不可思议的。

其次, 义务论反应倾向方面, 道德厌恶与义务论倾向的正相关关系也被许多研究所讨论(Chapman & Anderson, 2013; Haidt et al., 1997; Pizarro et al., 2011)。这类研究多认为, 厌恶刺激唤起决策者对道德规范违背行为的厌恶感, 特别是厌恶那些违背道德纯洁性的行为, 从而驱动更严厉的道德标准, 做出更义务论的道德判断(Haidt et al., 1997; Rozin et al., 1999; Wagemans et al., 2018)。另一方面, 积极情绪则可以抵消一定的厌恶感, 从而减少义务论反应或增加功利主义反应(Valdesolo & Desteno, 2006), 但也有研究指出快乐(Mirth)与升华感(Elevation)两种积极情绪, 其作用完全不同, 前者会降低义务论倾向, 而后者会增强义务论倾向(Strohminger et al., 2011)。因此, 情绪与义务论倾向的关系可能更加复杂, 在经典两难法的方法论框架下难以得到合理解释。

再次, 认知加工上的直觉/理性加工过程与行为反应上的义务论/功利主义反应是否是严格对应的关系? 从心理模型理论角度来看, 直觉功利主义是可能存在的(Bucciarelli, 2015), 近期研究使用二次回答范式也发现个体在直觉上可能就是功利主义的(Bago & de Neys, 2019a, 2019b), 个体在直觉加工状态下进行判断和在理性加工状态下进行再次判断, 其功利主义判断结果总体上没有变化。并且, 认知反省测试与功利主义反应之间并不具有显著的相关关系(Royzman et al., 2015)。因此, 直觉/理性加工与义务论/功利主义反应之间的对应关系也需要重新检验。甚至有批评者认为, 在两难框架下, 所谓的功利主义道德判断并不能反映人们对最大化结果的关注, 并进一步质疑经典两难法在揭示人们的道德决策规律方面的有效性(Kahane et al., 2015)。

经典两难法的最大局限性在于它的加工纯粹性假设, 它将决策者接受功利性提议与受到功利主义原则驱动划等号, 而将决策者不接受功利性提议与受到义务论原则驱动划等号。这使得义务论和功利主义原则严格对立, 互不相容, 功利主义倾向

越强，义务论倾向就会越弱，反之亦然。这与人们的常识不符，也与最近的研究发现相违背。常识上，面对一个道德选择时，人们完全有可能同时综合考虑行为背后的道德规范以及行为将会造成的道德后果。近期神经研究证据也表明，在道德决策过程中，前扣带回、脑岛和颞上回与情感评估相关，颞顶联合区和背内侧前额叶皮质与功利评估相关，而整体道德价值判断则表征在腹内侧前额皮质的前部；至关重要的是，这 3 组区域之间的反应和功能互动模式表明，情感和功利评估是独立并行计算的，并传递到腹内侧前额叶皮质，在那里，它们被整合成一个整体的道德价值判断(Hutcherson et al., 2015)。这意味着，功利主义和义务论反应倾向完全可以是相互独立，并行不悖的。

经典两难法的第二个局限性在于无法量化决策者在功利主义和义务论反应倾向程度上的差异。当这两种倾向程度相差很大时，决策者立马就可以做出判断，而当这两种倾向程度相差很小时，决策者需要更多的时间，但可能也做出了相同的判断。特别是在二元选择的情况下，决策要么接受功利性提议要么不接受，这种反应倾向的程度差异无法得到体现(Conway & Gawronski, 2013)。

经典两难法的第三个局限性在于其解释的模糊性(Gawronski & Beer, 2017)。该方法将功利主义原则与义务论原则严格对立，功利主义反应增多时(比如在类似电车难题中，选择赞成功利性提议的概率增加时)既可以解释为功利主义倾向强化的结果，也可以解释为义务论倾向弱化的结果，因此，在解释上存在模糊性。此外，决策者赞成功利性提议，既可能是功利主义原则驱动的结果，也可能只是一般性接受提议偏好的结果(比如电车难题中，决策者根本不管是否符合道德规范，也不管是否结果有利，只是一般性地偏好接受行为提议)。同理，拒绝功利性提议，既可能是义务论原则驱动的结果，也可能是偏好一般性不接受行为提议的结果。

为了解决前两个局限性，研究者提出了加工分离法(Process Dissociation, PD)，而为了解决前述所有局限性，发展出了 CNI 模型法(Gawronski et al., 2017)和 CAN 算法(Liu & Liao, 2021)。

3 加工分离法

加工分离法是 Conway 和 Gawronski (2013)从记忆研究中的加工分离程序(process dissociation procedure, Jacoby, 1991)发展出来的。Jacoby 最早使用 PD 方法来分离记忆成绩当中的回忆成分和基于熟悉性猜测的成分。这种方法并非内容特异性的，可以应用在许多领域当中(Hütter & Klauer, 2016)。Conway 和 Gawronski 第一次将它引入道德心理学的研究中以区分功利主义原则和义务论原则的驱动程度。

加工分离法实质上是操控行为提议所造成结果的利弊情况，当结果利大于弊时与牺牲性提议构成不一致情境，即虽然行为结果利大于弊但行为本身不符合道德规范，功利主义原则要求决策者接受该提议而义务论原则要求决策者拒绝该提议；而当结果弊大于利时则与伤害性提议构成一致情境，即行为结果弊大于利并且行为本身不符合道德规范，功利主义原则和义务论原则均要求决策者拒绝该提议。其原理如图 1 所示，不一致情境类似于经典两难情境，如前述的电车难题和天桥难题；一致情境将同一情境中行为提议所造成结果调整为弊大于利，比如电车难题中，失控的电车在原来轨道上行进只会轧死 1 名工人，若扳动道岔使其转入另一车道则会轧死 5 名工人，此时无论是受功利主义原则还是义务论原则驱动，这种伤害都是不可接受的；但若二者均未驱动，则人们可能会认为这种伤害也是可接受的。

在计算原理上，加工分离法使用了多个情境，每个情境均通过调整行为结果的利弊关系形成一致和不一致情境两个变式，以决策加工树的形式呈现出来。然后，通过计算决策者在一致情境和不一致情境中的选择伤害可接受或不可接受的概率来计算其功利主义和义务论倾向程度。因为一致情境和不一致情境条件下均有多个情境故事，因此，可以分别计算出决策者在两种条件下选择接受与不接受的反应概率。然后，使用这种反应概率推算出功利主义原则驱动程度与义务论原则驱动程度，具体计算如下：

$$p\,(\text{伤害不可接受}|\text{一致情境}) = U + (1-U)\times D \quad (1)$$

$$p\,(\text{伤害可接受}|\text{一致情境}) = (1-U)\times(1-D) \quad (2)$$

$$p\,(\text{伤害不可接受}|\text{不一致情境}) = (1-U)\times D \quad (3)$$

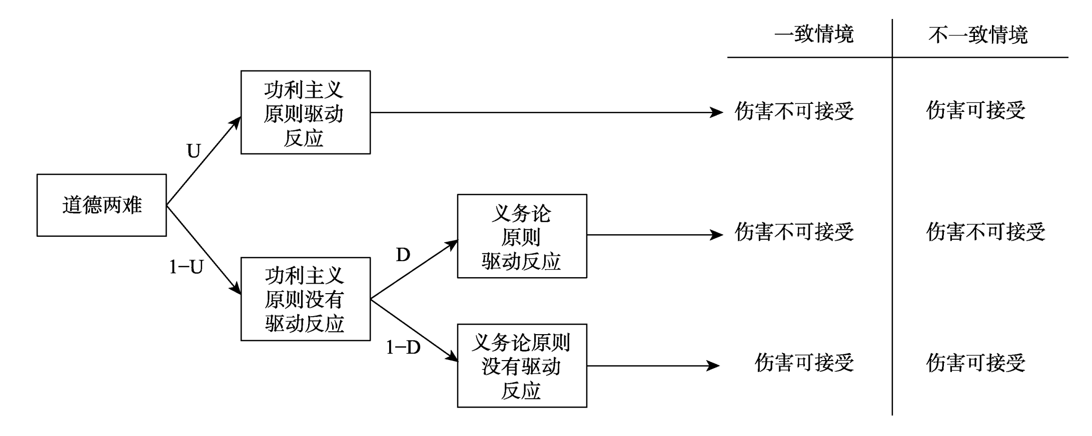


图 1　加工分离法所使用的决策加工树图解, 其中 U 为 Utilitarianism 的缩写, 代表功利主义原则驱动程度; D 为 Deontology 的缩写, 代表义务论原则的驱动程度。
资料来源：Conway & Gawronski (2013)。

p (伤害可接受|不一致情境) =

$$U + (1 - U) \times (1 - D) \quad (4)$$

(1)~(4)等式左侧是决策者在一致和不一致情境下接受和不接受行为提议的概率, 可以从决策者的决策结果直接计算出来。比如, 如果在不一致情境下有 8 个情境故事, 决策者在其中的 6 个故事中选择接受, 那么, p (伤害可接受|不一致情境) = 0.75, 而 p (伤害不可接受|不一致情境) = 0.25。同理, 也可以计算出一致情境下决策者接受和不接受行为提议的概率。然后使用等式(1)~(4)计算出 U 和 D 的值, 比如, 通过等式(4)减去等式(2), 或者通过等式(1)减去等式(3)可以得到:

U = p (伤害可接受|不一致情境) − p (伤害可接受|一致情境)

或

U = p (伤害不可接受|一致情境) − p (伤害不可接受|不一致情境)

计算出 U 的值之后, 再结合等式(1)~(4)当中的任何一个等式即可计算出 D 的值。

加工分离法相对于经典两难法, 在理论和方法上均往前推进了一步, 在经典两难法的单类型情境基础上扩展到双类型情境, 其理论价值主要在于:

首先, 加工分离法破除了经典两难法中功利主义反应(即接受功利性的行为提议)与受功利主义原则驱动和义务论反应(即拒绝功利性的行为提议)与受义务论原则驱动之间的一一对应关系。如此一来, 人们在做出某个选择时, 其背后的功利主义倾向与义务论倾向便都可以得到测量, 而并非互斥关系。

其次, 加工分离法引入一致性情境作为参照, 从而将功利主义原则和义务论原则都没有驱动行为反应的情形分离出来, 即图 1 最下面一行的情形。而这种情形在经典两难法中一直与功利主义倾向的表现相混淆。

Conway 和 Gawronski 在开发这一方法的同时, 进行了 3 个实证研究, 其结果进一步支持了道德双加工理论(Conway & Gawronski, 2013)。他们发现, 义务论倾向的确根植于对伤害行为的情绪反应, 在共情关心和观点采择上的个体差异与反应义务论倾向的 D 参数相关, 但与反应功利主义倾向的 U 参数不相关, 强化共情关心也只会影响 D 参数而不影响 U 参数; 认知需求的个体差异则与 U 参数相关, 与 D 参数不相关, 认知负荷的操控也只会选择性地影响 U 参数而并不影响 D 参数。

加工分离法对于在经典两难法研究中所得出的许多不一致的结论都给出了更多参考。比如, 有研究使用经典两难法发现男性比女性更加功利主义, 这在经典两难法视域下既可能是因为男性的功利主义倾向更强也可能是义务论倾向更弱, 而使用加工分离法则综合表明, 女性比男性的义务论倾向高, 男性比女性的功利主义倾向高, 但前者涉及中等程度差别, 而后者涉及差别较小。这说明道德两难问题上的性别差异主要是因为对伤害的情绪反应差异而不是对结果的认知评价差异所导致的(Friesdorf et al., 2015)。又如, 有学者

认为牺牲性判断所反应的是反社会性而并非真正的功利主义(Kahane et al., 2015), 而使用加工分离法则反映出反社会性可预测义务论倾向降低, 但不能预测功利主义倾向升高, 无论是哲学家还是普通人, 牺牲性功利主义判断都反映了真正的道德关怀(Conway et al., 2018)。该方法在目前学界应用非常广泛, 比如最近仍有研究者使用该方法发现推理能力和认知风格与功利性偏好存在正相关而与伤害性关怀则不相关, 从而进一步支持了道德双加工理论(Patil et al., 2021)。

但是, 加工分离法在革新经典两难法上并不彻底, 它只考虑了规范禁止时的两种情境, 而对于规范提倡时的情境则缺乏考察。禁止性规范与提倡性规范是道德规范的两个重要面向(Janoff-Bulman et al., 2009), 综合考察才能真正揭示规范在道德困境决策中的作用。功利主义要求最大化结果, 在研究中应该对结果的利弊差异进行控制; 而义务论强调行为要符合道德规范, 在研究中应该对行为提议是否符合道德规范进行控制。因此, 便形成了 4 种可能的组合情况: (1)行为提议的结果利大于弊, 但为规范所禁止; (2)行为提议的结果弊大于利, 且为规范所禁止; (3)行为提议的结果利大于弊, 且为规范所提倡; (4)行为提议的结果弊大于利, 但为规范所提倡。

尽管加工分离法在理论和方法上推进了道德困境研究, 但其理论框架仍不完善。在规范(禁止或提倡)与结果(利大于弊或弊大于利)的 4 种可能组合中, 只涵盖其中两种, 而缺乏对另外两种组合情形的考察。这导致两个方法学局限: 一是无法分离出决策者一般性地不接受或者接受行为提议的倾向。在实际决策行为中存在着这样一种决策行为, 它并不关心决策背后的规范或结果, 仅仅是一般性地对行为提议表示不接受或者接受, 这种倾向无法在加工分离法中得到呈现。二是如同经典两难法将一般性接受倾向与功利主义倾向相混淆一样, 加工分离法将一般性不接受倾向与义务论倾向相混淆。在图 1 的决策加工树第二行, 决策者在一致和不一致情境下均选择不接受行为提议, 这既可能是受义务论原则的驱动, 也可能是决策者在一般性不接受行为提议而并不顾及行为提议是否符合规范或行为结果是否有利。为了解决这一重要局限, Gawronski 研究团队于 2017 年使用计量经济学中的多项式决策加工树模型发展出了 CNI 模型法, 既分离出决策受功利主义(或者结果主义)原则驱动的程度和受义务论(或者规范主义)原则驱动的程度, 又分离出决策者不顾规范或结果而一般性不接受/接受倾向的程度。

4 CNI 模型法

CNI 模型法在加工分离法基础上, 将功利主义倾向操作化为对行为结果的敏感性, 将义务论倾向操作化为对行为规范的敏感性, 并进一步分离了一般性接受/不接受倾向, 在情境类型上涵盖了规范(禁止或提倡)与结果(利大于弊或弊大于利)的 4 种组合。多项式决策加工树模型是 CNI 模型法的基础, 在实证伦理学、社会心理学的许多领域中具有广泛应用(Hütter & Klauer, 2016; 刘媛媛 等, 2019)。CNI 模型法能够同时量化出决策者的结果敏感性(Sensitivity to Consequences), 规范敏感性(Sensitivity to Norms)和(不顾规范与结果的)一般性不接受或接受倾向(General preference for Inaction versus action irrespective of consequences and norms), 因此, 该方法被称之为道德决策的 CNI 模型法, 国内学者也介绍过该方法(刘媛媛 等, 2019; 徐科朋 等, 2020; 曾笑雨, 马燚娜, 2020)。

CNI 模型法假定决策者在进行道德决策时, 遵循序列加工规律, 使用决策加工树模型则可以将不同的心理加工过程逐层剥离出来, 如图 2 所示。当受功利主义原则驱动时, 决策者在结果有利时接受提议而在结果有弊时不接受提议(图 2 第一行); 如果没有受功利主义原则驱动而受义务论原则驱动, 决策者便会在规范提倡时接受提议而在规范禁止时不接受提议(图 2 第二行); 如果既没有受功利主义原则驱动, 也没有受义务论原则驱动, 决策者便会选择一般性地不接受(图 2 第三行)或者接受(图 2 第四行)行为提议。

在 CNI 模型法中, 决策者在不同条件下的决策结果所反映出的概率原理与加工分离法类似:

$$p\,(\text{接受}|\text{规范禁止, 利大于弊}) = C + (1 - C) \times (1 - N) \times (1 - I) \quad (5)$$

$$p\,(\text{接受}|\text{规范禁止, 弊大于利}) = (1 - C) \times (1 - N) \times (1 - I) \quad (6)$$

$$p\,(\text{接受}|\text{规范提倡, 利大于弊}) = C + (1 - C) \times N + (1 - C) \times (1 - N) \times (1 - I) \quad (7)$$

$$p\,(\text{接受}|\text{规范提倡, 弊大于利}) = (1 - C) \times N + (1 - C) \times (1 - N) \times (1 - I) \quad (8)$$

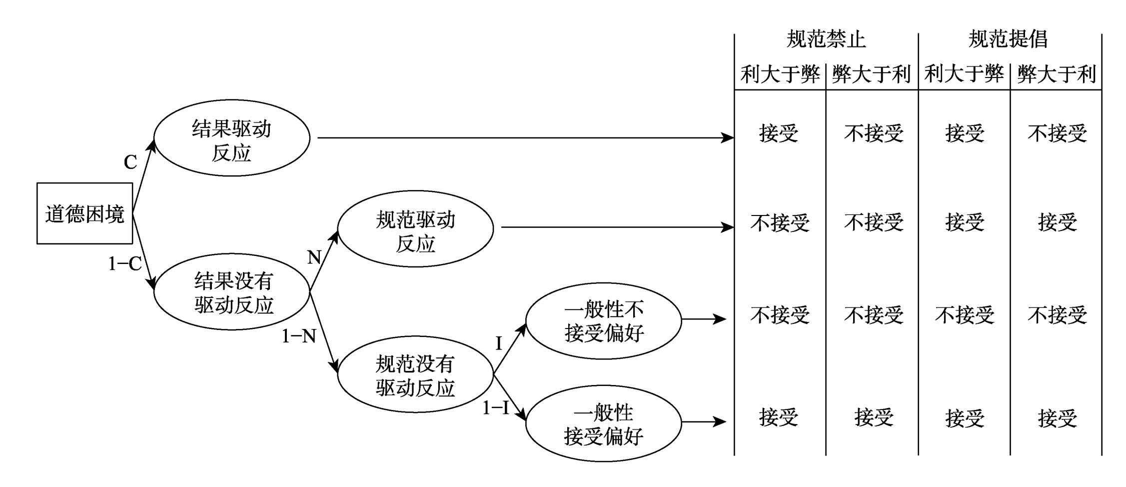


图 2　CNI 模型的多项式决策加工树图解, 其中 C 代表结果敏感性, N 代表规范敏感性, I 代表一般性不接受偏好(Gawronski 等认为一般性接受偏好的反应概率与一般性不接受偏好的反应概率之和为 1, 因此在模型中仅用一个 I 参数即可表达)。

资料来源：Gawronski et al. (2017).

由于在同类情境之下的多个情境故事中, 选择接受和不接受的概率之和为 1, 因此, 其概率等式在统计上具有等价意义, 此处只列了在 4 种组合情境下决策者接受提议的概率等式, 为了简化表达式, 依次使 *p* (接受|规范禁止, 利大于弊)为 *p*1, *p* (接受|规范禁止, 弊大于利)为 *p*2, *p* (接受|规范提倡, 利大于弊)为 *p*3, *p* (接受|规范提倡, 弊大于利)为 *p*4, 下文相同。

通过 Gawronski 提供的决策加工树工具可以直接计算出决策者的结果敏感性(C 参数)、规范敏感性(N 参数)和一般性不接受/接受倾向(I 参数) (http://www.bertramgawronski.com/documents/CNI-Model_Materials.zip), 同时, 也可以对这 3 个参数在两组数据之间进行比较或者与某个特定值进行比较。所得到的 C 和 N 参数如果显著大于 0 则表明决策者有显著的结果和规范敏感性, 如果 I 参数显著大于 0.5 则表明决策者有显著的一般性不接受倾向, 如果显著小于 0.5 则表明决策者有显著的一般性接受倾向。

前已述及, 为了解决经典两难和加工分离法所存在的局限性, Gawronski 团队发展出了 CNI 模型法(Gawronski et al., 2017; Gawronski & Beer, 2017)。这一方法对于解决道德困境研究中的争议问题具有促进作用。比如, 前人发现情绪对于道德判断具有重要作用, 但不能确定情绪作用于规范敏感性还是结果敏感性, 抑或只是作用于一般性反应倾向, 而使用 CNI 模型法则表明, 启动高兴情绪可以降低对道德规范的敏感性, 而没有影响对结果的敏感性和一般性不接受或接受倾向; 启动悲伤和愤怒则对道德判断没有显著影响(Gawronski et al., 2018)。前人有研究表明睾丸酮素增强决策者的功利主义反应倾向, 而使用 CNI 模型法则发现, 外源性类固醇睾丸酮素使决策者对规范的敏感性增强; 相反, 内源性睾酮在基线测量时的模式正好相反, 内源性睾酮水平越高, 对道德规范的敏感度越低。研究结果表明, 睾丸激素在道德判断中的作用比之前的研究结果更为复杂(Brannon et al., 2019)。近来, Gawronski 团队还使用 CNI 模型法进一步探析了道德两难决策中的外语效应(Białek et al., 2019)、权力效应(Gawronski & Brannon, 2020)、政治意识形态与道德决策之间的关系(Luke & Gawronski, 2021a)、精神病性与道德决策之间的关系(Luke & Gawronski, 2021b)等。国内学者李中权等则运用该方法探讨了压力与道德决策之间的关系(Li et al., 2019; Zhang et al., 2018)。在 CNI 模型法基础上, Hennig 和 Hütter (2020)开发了升级版 proCNI 模型法来重新审视功利主义与义务论之间的区别。他们发现二者的区分是人为的, 人们对这两种原则的敏感性是相互关联的。这些研究均展现了 CNI 模型这样的形式化建模方法对于深入探讨道德判断影响因素的方法学价值。

CNI 模型法在理论构念上全面覆盖了规范(禁止或提倡)与结果(利大于弊或弊大于利)的 4 种组合情况，在方法论上具有突破性的贡献，特别是对规范禁止和规范提倡两类情境的综合考察，对常人道德世界的刻画更为准确。有学者指出，常人的道德是“一种合宜恰适的尺度，其行为基准既不应低于这一尺度也无须高于这一尺度” (甘绍平, 2017)。这里的尺度，即是规范禁止与规范提倡之间的道德空间。尽管 CNI 模型法具有理论构念上的突破性，这一方法依然存在着一些局限性：

首先，CNI 模型法最大的问题在于其理论逻辑先验地假定了决策者的决策过程是一种特定的序列加工过程：决策者首先考虑的是结果，在不考虑结果的情况下再考虑规范，在结果与规范均不考虑的情况下，才会表现出一般性的不接受或者接受倾向。CNI 模型法的构念逻辑是逐层剥离，而非平行建构的。而实际上，决策者完全有可能是并行加工进行决策，即存在 3 种可能性：在决策时同时考虑决策背后的规范以及决策可能导致的结果，抑或先考虑决策是否符合规范然后再考虑决策的结果，也有可能首先只是形成一般性行为态度，然后受到规范或结果原则的修正。因此，倘若将图 2 当中各参数的位置互换之后，其算法结果的概率模型将完全不同，下文详述。

其次，该方法不能应用于相关或回归研究设计。CNI 模型法的底层是多项式决策加工树模型，这种方法广泛应用于区分前置变量中的多重影响因素。但是，CNI 模型法是以群体比较而非个体比较为基础的，所生成的 C、N、I 参数是表征在群体层面而非个体层面的，因此不能进行相关或回归分析。

再次，受限于多项式加工决策树工具，CNI 模型法只能比较两组参数之间或者某个参数与参照值之间的差异性检验，而不能进行多组数据之间的比较。因此，其计算方法的拓展性较为受限。

CNI 模型法的后两方面局限性已在近日得到一定程度的解决，Gawronski 团队通过扩展情境数量和提供个体层面的参数计算方案，使得 C、N、I 参数可用于相关、回归和多重比较等个体差异的研究(Korner et al., 2020)。但是，其最大局限——序列加工的前提假设仍然未得到解决，CAN 算法则对以上局限进行了修正。

5 CAN 算法

前已述及，CNI 模型法虽有突破性的贡献，但也存在着局限性。特别是在理论构念上，它假定决策者首先考虑行为结果，然后再考虑行为的规范性，最后才考虑一般性行为倾向的梯级序列。这一构念上的基本假设没有依据，也受到学者们的批评(Baron & Goodwin, 2020; Liu & Liao, 2021), Gawronski 在回应中也承认这一顺序假设可能会造成参数估计时的偏差，但否认其对于整体参数解释具有影响，认为这只是一种逻辑上的条件关系而非时间上的序列关系(Gawronski et al., 2020)。

如果序列加工的决策模式对于量化决策者的各个心理过程没有影响，那么，决策者除了按 CNI 模型法所假定的结果→规范→一般性不接受/接受倾向序列进行决策之外，也可能按规范→结果→一般性不接受/接受倾向，或者一般性不接受/接受倾向→受规范修正→受结果修正，或者一般性不接受/接受倾向→受结果修正→受规范修正等序列加工思路进行决策。试以规范→结果→一般性不接受/接受倾向这一序列为例(图 3)。

相应地，决策者在各类情境下的决策反应概率为：

$$p1 = (1-\mathrm{N})\times\mathrm{C} + (1-\mathrm{N})\times(1-\mathrm{C})\times(1-\mathrm{I}) \quad (9)$$

$$p2 = (1-\mathrm{N})\times(1-\mathrm{C})\times(1-\mathrm{I}) \quad (10)$$

$$p3 = \mathrm{N} + (1-\mathrm{N})\times\mathrm{C} + (1-\mathrm{N})\times(1-\mathrm{C})\times(1-\mathrm{I}) \quad (11)$$

$$p4 = \mathrm{N} + (1-\mathrm{N})\times(1-\mathrm{C})\times(1-\mathrm{I}) \quad (12)$$

倘若决策加工序列对其决策倾向的估计是没有影响的，那么，等式(5)~(8)与等式(9)~(12)在统计上应当是分别等价对应的。然而，稍加换算便可发现这种情形在统计上成立的可能性极小，比如将图 2 模型的概率等式中 N 参数经换算后得到 $\mathrm{N} = (-p1 - p2 + p3 + p4)/(2 - p1 + p2 - p3 + p4)$；而将图 3 模型的概率等式中 N 参数经换算后得到 $\mathrm{N} = (-p1 - p2 + p3 + p4)/2$。如果 CNI 模型中的参数位置关系并不会影响概率估计结果的话，这两个 N 参数应当相等。如此，则会换算出 $p2 - p1 = p3 - p4$，这在经验统计上成立的概率极小。Gawronski 等(2017)在脚注 7 里报告称，当 C 和 N 的位置互换之后，可以重复出他们研究的效应结果，但对于部分差异边缘显著的结果就会变得显著，因此，他们未对此更进一步分析，也未说明

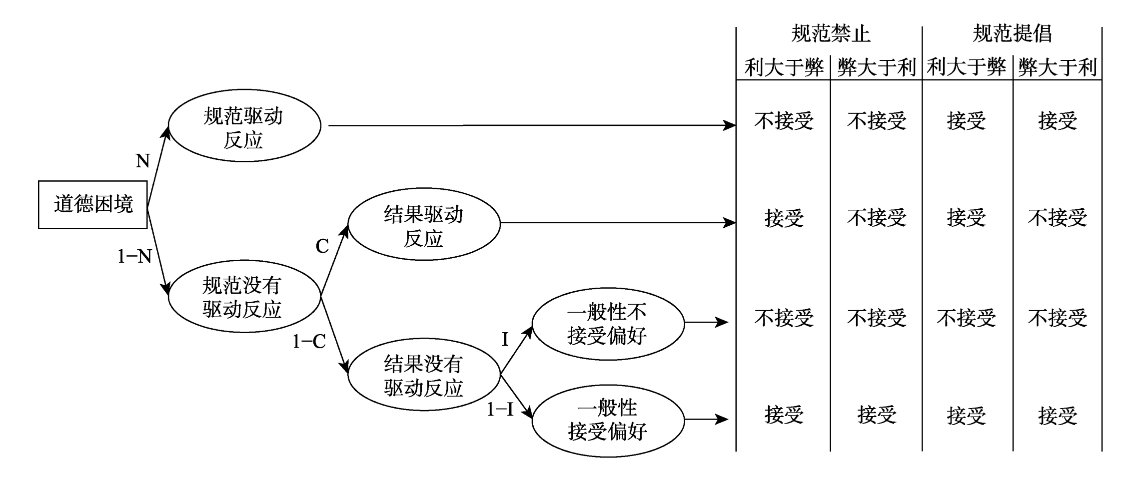


图 3　规范→结果→一般性不接受/接受倾向下的决策加工树模型
资料来源：Liu & Liao (2021)。

为何会出现统计变得更显著的原因。但是，这恰恰也说明 C 和 N 的位置对于最终的参数估计结果存在影响，只是这种影响恰好符合作者的假设便未得到进一步辨析。

此外，Gawronski 等(2017)认为一般性行为反应倾向只能在模型的最低层级，因为他们的反应概率之和为 1，是非此即彼的矛盾反应倾向。这一先验假设也未必成立。决策者完全可能首先拥有一般性接受或不接受反应倾向，而后受到规范或结果原则的修正而改变了原来的反应倾向，即在图 2 和图 3 当中 C, N, I 三个参数之间的位置关系在逻辑上是不确定和可以互换的。

从以上分析可以得出，决策者的决策模式如果是序列加工形态的话，应当存在着多种决策模式，而 CNI 的模型算法显然只是针对其中一种。此外，决策者更加可能会同时对规范与结果进行权衡，才会出现冲突性的两难。如果决策者只是序列性地选择其中某一条决策路径，则在实际决策时不会出现规范与结果相冲突的情形。因此，在存在潜在两难的问题中，决策加工过程更可能是并行而非序列加工形态，Liu 和 Liao (2021)则对此进行了深入辨析。

为了解决上述局限性，研究者在 CNI 模型法的基础上发展出了 CAN 算法。比如在 CNI 模型的情境结构中，在 4 种组合情况下均有 6 个情境故事，因此，从经验概率上可以计算出每种组合情况下个体在 6 个情境故事中选择接受行为提议的概率。结合这 4 个概率数据，综合考虑决策者可能进行并行加工的决策特点，也可以合成与 CNI 模型类似的 3 个维度指标。为了与 CNI 模型法进行区分并方便记忆，称之为 CAN 算法(Liu & Liao, 2021)：

结果敏感性指标 C (Consequence sensitivity, C) = $1/2(p1 - p2 + p3 - p4)$;

规范敏感性指标 N (Norm sensitivity, N) = $1/2(p3 - p1 + p4 - p2)$;

总体接受/不接受指标 A (overall Action/Inaction preference, A) = $1/4(p1 + p2 + p3 + p4)$。

这种合成方法常见于文献当中，比如 Talhelm 等(2014)计算忠诚/裙带关系(loyalty/nepotism)指标时，用参与者奖励朋友的数量减去惩罚朋友的数量。这与 CAN 算法使用规范提倡时对行为提议的赞成概率减去规范禁止时对行为提议的赞成概率来表示规范驱动指标，以及使用结果利大于弊时对行为提议的接受概率减去结果弊大于利时对行为提议的接受概率来表示结果驱动指标，在原理上是相通的。

CAN 算法从并行加工角度假定决策者同时考虑到规范与结果两个方面，并且受二者交互影响。如果规范对决策者具有正面影响，那么，无论结果利大于弊或者弊大于利时，规范提倡情况下决策者对行为提议的接受概率应当显著大于规范禁止时的接受概率，二者之差即可代表规范的敏感程度。当利大于弊时，规范的敏感程度可以用 $p3 - p1$ 来表示，当弊大于利时，规范的敏感程度可以用 $p4 - p2$ 来表示。那么，对规范的整体敏感

程度也就可以用二者的平均数来表示，即 1/2($p3$ − $p1$ + $p4$ − $p2$)。同理，如果结果对决策者具有正面影响，那么，无论规范提倡或者禁止时，结果利大于弊的情况下决策者对行为提议的接受概率应当显著大于弊大于利时的接受概率，二者之差即可代表结果的敏感程度。当规范禁止时，结果的敏感程度可以用 $p1$ − $p2$ 来表示，当规范提倡时，结果的敏感程度可以用 $p3$ − $p4$ 来表示。那么结果的整体敏感程度也就可以用二者的平均数来表示，即 1/2($p1$ − $p2$ + $p3$ − $p4$)。而作为整体反应倾向指标而言，所有情形下的平均接受概率则可以反映决策者的整体偏好性，即 1/4($p1$ + $p2$ + $p3$ + $p4$)。

例如，假设在规范与结果的 4 种组合情境下，个体选择接受行为提议的概率分别为 $p1$ = 0.6, $p2$ = 0.1, $p3$ = 0.9, $p4$ = 0.5。那么，其规范敏感度 1/2($p3$ − $p1$ + $p4$ − $p2$) = 0.35；其结果敏感性 1/2($p1$ − $p2$ + $p3$ − $p4$) = 0.45；其整体偏好性 1/4($p1$ + $p2$ + $p3$ + $p4$) = 0.525。同理，即可计算出每一个个体在这 3 个指标上的值，然后进行统计检验。

在 CAN 算法的指标解读上，C 指标如果在统计上显著大于 0，这说明与行为提议造成结果弊大于利的情况相比，决策者在利大于弊的情况下对提议的接受概率更高，因此受到显著的功利主义(或者结果主义)原则驱动；如果显著小于 0，说明决策者受到显著的反功利主义(或者反结果主义)原则驱动；如果与 0 无显著差异，说明决策未受到功利主义原则的驱动。同理，N 指标如果显著大于 0，说明决策者受到显著的义务论(或者规范主义)原则驱动；如果显著小于 0，说明决策者受到显著的反义务论(或者反规范主义)原则驱动；如果与 0 无显著差异，说明决策者未受到义务论原则的驱动。对于 A 指标，如果显著大于 0.5，说明其具有总体接受倾向相比于不接受倾向的优势；如果显著小于 0.5，说明其具有总体不接受倾向相比于接受倾向的优势；如果与 0.5 无显著差异时有两种情况：1、如果 C 指标和 N 指标同时与 0 也无显著差异，说明决策者只是完全随机作答；2、如果至少 C 和 N 指标其中之一与 0 有显著差异，说明决策者总体接受与不接受倾向强度相当，但也受到了功利主义或义务论原则的影响。

CAN 算法是建立在 CNI 模型法基础上的，全面汲取了 CNI 模型法的理论构念优势，但二者在 5 个方面存在差异：第一，CNI 模型法假定决策者在进行道德决策时是基于先结果后规范再一般性不接受/接受偏好的序列加工模式，而 CAN 算法则没有这样的先验假设，而使用一个常见的参数计算中的代数算法，将规范与结果的作用同等看待，取其平均。第二，CNI 模型法最初的版本导出的参数是在群体层面的，因此不能用于相关或回归设计，参数也只能在两组之间比较或者与某一特定值进行比较，而 CAN 算法所得到的参数是在个体层面的，既可以用于相关或回归设计，也可以在多组之间进行比较。第三，CNI 模型法中 I 参数与 CAN 算法中的 A 参数在统计上的解释方向是相反的，前者数值越大代表一般性不接受倾向越强，后者数值越大代表总体接受倾向越强。第四，CNI 模型中的 I 参数是以序列加工为假设前提，在剥离结果和规范的作用之后，代表决策者一般性不接受/接受倾向的程度，而 CAN 算法中的 A 参数是综合考虑结果或规范作用在内，代表决策者总体上接受/不接受倾向的程度，其内涵是不同的。第五，CNI 模型法依赖于二元反应模式，个体必须对行为提议做二元反应，要么接受要么不接受，而 CAN 算法无此要求，也可用于连续性评分设计，决策者通过对行为提议的接受或不接受程度进行评分也可以实现相应的参数计算。综合起来，CNI 模型法和 CAN 算法，在实际研究中可以综合运用，相互参照，特别是 CAN 算法，已在道德决策的相关研究中得到了应用(Liu & Liao, 2021；刘传军，廖江群，印刷中)。

在研究应用中，CAN 算法的局限性也应当受到重视。首先是测量误差问题，当运用 CAN 算法的指标与其他心理变量进行相关分析时，可能会因为存在测量误差导致相关显著或不显著。目前 CAN 算法还没有统计指标来显示这种测量误差的程度以及校正方法，有待后续研究继续探索。其次，与 CNI 模型法一样，CAN 算法也只有 6 个情境共 24 个试次，而观察数偏少可能存在一定测量变异。该局限可以通过增加情境试次数量或进行分时段多次重复测量来平衡这种变异。

6 讨论与展望

道德困境研究是社会困境研究中的典型代表，其研究方法对于其他社会困境研究具有参考性和可迁移性。纵览道德困境研究的 4 种实证方法，从

经典两难法和加工分离法，到 CNI 模型法和 CAN 算法，在理论构念上是向下兼容的，越往后的方法涵盖了之前方法的理论构念。特别是 CNI 模型法和 CAN 算法，实现了对规范(禁止或提倡)与结果(利大于弊或弊大于利)的全组合情境的通盘考察，在未来的道德决策研究当中应该重点加以应用，在其他社会困境研究中也可以迁移使用。

CNI 模型法和 CAN 算法对于以往悬而未决的许多争议均具有很强的指导意义。首先，以往研究中所发现的精神病性与功利主义反应的正相关关系，可能只是因为精神病性与一般性或总体不接受/接受倾向存在相关，而与结果敏感性未必存在着关联，这需要后续研究进一步检验。其次，前人研究表明厌恶感会使道德判断更为严厉(Haidt et al., 1997; Rozin et al., 1999; Wagemans et al., 2018)，但近期研究表明并无此效应(Johnson et al., 2016)。这可能是由于使用经典道德两难法的方法学缺陷所致，可使用 CNI 模型和 CAN 算法来进一步检验。此外，情绪性厌恶唤醒究竟是强化了个体的规范敏感性，还是只是增强了其一般性或总体不接受倾向，也可以通过本文中的后两种方法得到解答。再次，个体认知的双加工状态与功利主义/义务论倾向之间是否具有一一对应关系，也需要重新检验，有学者提出存在直觉功利主义的可能(Bago & de Neys, 2019a, 2019b)，这也可以使用后两种方法来进行检验。更进一步，有学者使用 CNI 模型的变体来考察规范/结果敏感性与义务论/功利主义的对应关系，发现代表规范的参数对结果也很敏感，由此得出，除非期望规范产生切实的结果，否则它们不会指导道德判断，这表明规范和结果(或义务论和功利主义)作为道德判断的决定因素之间的分裂是人为的(Hennig & Hütter, 2020)。实际上这也就是 CAN 算法当中，存在着规范敏感性与结果敏感性的交互作用。其他的许多研究争议也可以在 CNI 模型法和 CAN 算法的视域下得到更深入解答。

CNI 模型法和 CAN 算法还具有非常强大的理论和方法拓展性。这种方法作为一种理论和方法思路，并不仅仅只能用于道德困境的研究，也可以应用于其他具有潜在冲突性的研究议题上，试举 3 例。在确定行为的道德可责性上一直存在着意图与后果的争议，意图论者认为行为的道德可责性在于其意图是否存在主观故意，而后果论者认为行为的道德可责性在于其后果是否存在伤害或妨碍性。那么，对于确定行为的道德可责性，意图与后果便形成潜在的冲突，可能存在故意而有害、故意而无害、非故意而有害和非故意而无害等情境组合，研究者可以结合 CNI 模型法和 CAN 算法的理论构念对此进行研究。近期有学者在研究对伤害的情绪反应时，也推荐了这种方法学思路(Reynolds & Conway, 2018)。同理，在个人道德偏好上，公义与私利也常常形成潜在冲突，出现益公益私、益公害私、害公益私和害公害私等情境组合，研究者也可以通过类似的方法进行剖析。又如，在消费心理学研究当中，人们对目标商品的评价可能受到在时空上接近的其他非目标商品的情感效价的影响，进而影响其消费决策。在这种评价性条件效应当中可能存在着可控和不可控的认知加工成分，研究者也可以通过类似的思路进行深入探讨(Hütter et al., 2018)。再如，员工的亲组织不道德行为(Umphress et al., 2010)，具有亲组织性和不道德性的冲突性成分，导致员工既会感受到骄傲也会感受到内疚的情绪(Tang et al., 2020)。这也可以使用本文中的研究方法来对其亲组织性心理成分和不道德性心理成分进行独立测量。因此，CNI 模型法和 CAN 算法，既可应用于探讨道德困境研究中除规范与结果冲突以外的其他冲突成分(Everett & Kahane, 2020)，也可迁移应用于其他领域，只要具有潜在冲突性的研究议题都可以考虑使用与之类似的方法来进行探讨。

综上，道德困境研究走过了经典两难法、加工分离法、CNI 模型法和 CAN 算法 4 种方法学范式的沿革，在理论构念上从单类型冲突情境，向规范(禁止或提倡)与结果(利大于弊或弊大于利)的 4 种全组合析因情境发展。特别是 CNI 模型法和 CAN 算法，对于解决以往研究中的许多争议提供了方法学思路。并且，这种思路框架并不局限于道德困境研究，也为其他具有潜在冲突性的社会困境和其他议题提供了方法学借鉴。

参考文献

甘绍平. (2017). 常人道德的尺度. *道德与文明*, (3), 31–37.

刘传军，廖江群. (印刷中). 不幸的道德：运气越差越功利主义. *中国社会心理学评论*.

刘媛媛，丁一，彭凯平，胡传鹏. (2019). 多项式加工树模

型在社会心理学中的应用. *心理科学, 42*(2), 422–429.

徐科朋, 杨凌倩, 吴家虹, 薛宏, 张姝玥. (2020). CNI 模型在道德决策研究中的应用. *心理科学进展, 28*(12), 2102–2113.

喻丰, 彭凯平, 韩婷婷, 柴方圆, 柏阳. (2011). 道德困境之困境——情与理的辩争. *心理科学进展, 19*(11), 1702–1712.

曾笑雨, 马燚娜. (2020). 多项式模型在道德判断研究中的应用. *科学通报, 65*(19), 1912–1921.

Bago, B., & de Neys, W. (2019a). Advancing the specification of dual process models of higher cognition: A critical test of the hybrid model view. *Thinking & Reasoning, 26*(1), 1–30.

Bago, B., & de Neys, W. (2019b). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General, 148*(10), 1782–1801.

Baron, J., & Goodwin, G. P. (2020). Consequences, norms, and inaction: A critical analysis. *Judgement and Decision Making, 15*(3), 421–442.

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*(1), 154–161.

Bentham, J. (1996). *An Introduction to the Principles of Morals and Legislation*. New York: Oxford University Press, USA. (Original work published 1781).

Białek, M., Paruzel-Czachura, M., & Gawronski, B. (2019). Foreign language effects on moral dilemma judgments: An analysis using the CNI model. *Journal of Experimental Social Psychology, 85*, 103855.

Bonnefon, J. -F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science, 352*(6293), 1573–1576.

Brannon, S. M., Carr, S., Jin, E. S., Josephs, R. A., & Gawronski, B. (2019). Exogenous testosterone increases sensitivity to moral norms in moral dilemma judgements. *Nature Human Behaviour, 3*(8), 856–866.

Bucciarelli, M. (2015). Moral dilemmas in females: Children are more utilitarian than adults. *Frontiers in Psychology, 6*, 1345.

Chapman, H. A., & Anderson, A. K. (2013). Things rank and gross in nature: A review and synthesis of moral disgust. *Psychological Bulletin, 139*(2), 300–327.

Christensen, J. F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience and Biobehavioral Reviews, 36*(4), 1249–1264.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology, 104*(2), 216–235.

Conway, P., Goldstein-Greenwood, J., Polacek, D., & Greene, J. D. (2018). Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition, 179*, 241–265.

Everett, J. A. C., & Kahane, G. (2020). Switching tracks? Towards a multidimensional model of utilitarian psychology. *Trends in Cognitive Science, 24*(2), 124–134.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 2*(2), 152–161.

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in responses to moral dilemmas: A process dissociation analysis. *Personality and Social Psychology Bulletin, 41*(5), 696–713.

Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Sposta, S., Zago, S., … Priori, A. (2010). Gender-related differences in moral judgments. *Cognitive Processing, 11*(3), 219–226.

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of Personality and Social Psychology, 113*(3), 343–376.

Gawronski, B., & Beer, J. S. (2017). What makes moral dilemma judgments "utilitarian" or "deontological"? *Social Neuroscience, 12*(6), 626–632.

Gawronski, B., & Brannon, S. M. (2020). Power and moral dilemma judgments: Distinct effects of memory recall versus social roles. *Journal of Experimental Social Psychology, 86*, 103908.

Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R., & Hütter, M. (2018). Effects of incidental emotions on moral dilemma judgments: An analysis using the CNI model. *Emotion, 18*(7), 989–1008.

Gawronski, B., Conway, P., Hütter, M., Luke, D. M., Armstrong, J., & Friesdorf, R. (2020). On the validity of the CNI model of moral decision-making: Reply to baron and goodwin (2020). *Judgment and Decision Making, 15*(6), 1054–1072.

Gold, N., Colman, A. M., & Pulford, B. D. (2014). Cultural differences in responses to real-life and hypothetical trolley problems. *Judgment and Decision Making, 9*(1), 65–76.

Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology, 8*, 125–130.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences, 11*(8), 322–323; author reply 323–324.

Greene, J. D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental*

Social Psychology, 45(3), 581–584.

Greene, J. D. (2016). Our driverless dilemma: When should your car be willing to kill you? *Science, 352*(6293), 1514–1515.

Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*(12), 517–523.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107*(3), 1144–1154.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*(5537), 2105–2108.

Gubbins, E., & Byrne, R. M. J. (2014). Dual processes of emotion and reason in judgments about moral dilemmas. *Thinking & Reasoning, 20*(2), 245–268.

Haidt, J., Rozin, P., Mccauley, C., & Imada, S. (1997). Body, psyche, and culture: The relationship between disgust and morality. *Psychology & Developing Societies, 9*(9), 107–131.

Hennig, M., & Hütter, M. (2020). Revisiting the divide between deontology and utilitarianism in moral dilemma judgment: A multinomial modeling approach. *Journal of Personality and Social Psychology, 118*(1), 22–56.

Hutcherson, C. A., Montaser-Kouhsari, L., Woodward, J., & Rangel, A. (2015). Emotional and utilitarian appraisals of moral dilemmas are encoded in separate areas and integrated in ventromedial prefrontal cortex. *Journal of Neuroscience, 35*(36), 12593–12605.

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology, 27*(1), 116–159.

Hütter, M., Sweldens, S., Morwitz, V., & Andrade, E. (2018). Dissociating controllable and uncontrollable effects of affective stimuli on attitudes and consumption. *Journal of Consumer Research, 45*(2), 320–349.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513–541.

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology, 96*(3), 521–537.

Johnson, D. J., Wortman, J., Cheung, F., Hein, M., Lucas, R. E., Donnellan, M. B., … Narr, R. K. (2016). The effects of disgust on moral judgments. *Social Psychological and Personality Science, 7*(7), 640–647.

Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition, 134*, 193–209.

Kant, I., & Gregor, M. J. (1997). *Groundwork of the metaphysics of morals*. Cambridge, UK: Cambridge University Press.

Korner, A., Deutsch, R., & Gawronski, B. (2020). Using the CNI model to investigate individual differences in moral dilemma judgments. *Personality and Social Psychology Bulletin, 46*(9), 1392–1407.

Liu, C., & Liao, J. (2021). CAN algorithm: An individual level approach to identify consequences and norms sensitivities and overall action/inaction preferences in moral decision-making. *Frontiers in Psychology, 11*, 547916.

Li, Z., Gao, L., Zhao, X., & Li, B. (2019). Deconfounding the effects of acute stress on abstract moral dilemma judgment. *Current Psychology*. Advance publish online https://doi.org/10.1007/s12144-019-00453-0

Lucas, B. J., & Livingston, R. W. (2014). Feeling socially connected increases utilitarian choices in moral dilemmas. *Journal of Experimental Social Psychology, 53*(5), 1–4.

Luke, D. M., & Gawronski, B. (2021a). Political ideology and moral dilemma judgments: An analysis using the CNI model. *Personality and Social Psychology Bulletin*. Advance publish online https://doi.org/10.1177/0146167220987990

Luke, D. M., & Gawronski, B. (2021b). Psychopathy and moral dilemma judgments: A CNI model analysis of personal and perceived societal standards. *Social Cognition, 39*(1), 41–58.

Mill, J. S. (1872). *The logic of the moral sciences*. Chicago, Illinois, US: Open Court.

Patil, I., Zucchelli, M. M., Kool, W., Campbell, S., Fornasier, F., Calo, M., … Cushman, F. (2021). Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures. *Journal of Personality and Social Psychology, 120*(2), 443–460.

Pizarro, D., Inbar, Y., & Helion, C. (2011). On disgust and moral judgment. *Emotion Review, 3*(3), 267–268.

Reynolds, C. J., & Conway, P. (2018). Not just bad actions: Affective concern for bad outcomes contributes to moral condemnation of harm in moral dilemmas. *Emotion, 18*(7), 1009–1023.

Royzman, E. B., Landy, J. F., & Leeman, R. F. (2015). Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. *Cognitive Science, 39*(2), 325–352.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76*(4), 574–586.

Seara-Cardoso, A., Dolberg, H., Neumann, C., Roiser, J. P., & Viding, E. (2013). Empathy, morality and psychopathic traits in women. *Personality and Individual Differences, 55*(3), 328–333.

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: A virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience, 8*, 426.

Strohminger, N., Lewis, R. L., & Meyer, D. E. (2011). Divergent effects of different positive emotions on moral judgment. *Cognition, 119*(2), 295–300.

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within china explained by rice versus wheat agriculture. *Science, 344*(6184), 603–608.

Tang, P. M., Yam, K. C., & Koopman, J. (2020). Feeling proud but guilty? Unpacking the paradoxical nature of unethical pro-organizational behavior. *Organizational Behavior and Human Decision Processes, 160*, 68–86.

Tassy, S., Deruelle, C., Mancini, J., Leistedt, S., & Wicker, B. (2013). High levels of psychopathic traits alters moral choice but not moral judgment. *Frontiers in Human Neuroscience, 7*(4), 229.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *Monist, 59*(2), 204–217.

Umphress, E. E., Bingham, J. B., & Mitchell, M. S. (2010). Unethical behavior in the name of the company: The moderating effect of organizational identification and positive reciprocity beliefs on unethical pro-organizational behavior. *Journal of Applied Psychology, 95*(4), 769–780.

Valdesolo, P., & Desteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17*(6), 476–477.

Wagemans, F. M. A., Brandt, M. J., & Zeelenberg, M. (2018). Disgust sensitivity is primarily associated with purity-based moral judgments. *Emotion, 18*(2), 277–289.

Young, A. D., & Monroe, A. E. (2019). Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. *Journal of Experimental Social Psychology, 85*, 103870.

Youssef, F. F., Dookeeram, K., Basdeo, V., Francis, E., Doman, M., Mamed, D., … Legall, G. (2012). Stress alters personal moral decision making. *Psychoneuroendocrinology, 37*(4), 491–498.

Yu, G., & Tang, S. (2013). Psychopathic personality and utilitarian moral judgment in college students. *Journal of Criminal Justice, 41*(5), 342–349.

Zhang, L., Kong, M., Li, Z., Zhao, X., & Gao, L. (2018). Chronic stress and moral decision-making: An exploration with the CNI model. *Frontiers in Psychology, 9*, 1702.

The development of empirical paradigms and their theoretical values in moral dilemma research

LIU Chuanjun[1,2,3], LIAO Jiangqun[3]

([1] *Department of Sociology and Psychology, School of Public Administration, Sichuan University, Chengdu* 610065, *China*)

([2] *Institute of Psychology, Sichuan University, Chengdu* 610065, *China*)

([3] *Department of Psychology, School of Social Sciences, Tsinghua University, Beijing* 100084, *China*)

Abstract: The present paper systematically reviewed the development procedure of moral dilemma research paradigms. Specifically, we discussed the advantages/disadvantages and the theoretical values of the four empirical paradigms: classical dilemma paradigm, process dissociation paradigm; CNI (Consequences, Norms, and generalized Inaction/action preferences) model and CAN (Consequences sensitivity, overall Action/inaction preferences and Norms sensitivity) algorithm. The later paradigms resolved the limitations of the former. The process dissociation paradigm overcame the limitations of classical moral dilemma paradigm, such as the pure process hypothesis. The CNI model further dissociated multiple psychological processes of moral dilemma decisions on the basis of process dissociation paradigm. CAN algorithm addressed the improper presuppose of sequential process in the CNI model. Future researchers can apply the newest method to solve the controversies in empirical findings and reinspect the existing moral theories. They can apply relevant methods to explore other potentially conflicting research issues. To summarize, the present paper provides a methodological reference for moral dilemmas and related research.

Key words: moral dilemma, process dissociation, CNI model, CAN algorithm, moral decision-making